Anna Maria Buehler[1], Alexandre Biasi Cavalcanti[1],Erica Aranha Suzumura[1], Mariana Teixeira Carballo[1], Otávio Berwanger[1]

How to assess intensive care randomized trials

Como avaliar criticamente um ensaio clínico de alocação aleatória em terapia intensiva

1. Instituto de Ensino e Pesquisa do Hospital do Coração - IEP-HCor - São Paulo (SP), Brazil.

ABSTRACT

Randomized controlled trials are scientific experiments considered the gold-standard to evaluate therapeutic interventions. Randomized controlled trials may examine the safety and efficacy of new drugs and therapeutic procedures or compare the effects of two or more drugs or any other intervention. In this article, we present the essential features of these studies, as well as factors which may bias randomized controlled trials. We also present criteria to critically appraise articles reporting randomized controlled trials, explain how to interpret results and how to apply them to clinical practice.

Keywords: Evaluation; Randomized controlled trials as topic; Evidence-based medicine/methods; Intensive care units

Author for correspondence:
Otávio Berwanger
Instituto de Ensino e Pesquisa do Hospital do Coração (IEP-HCor)
Rua Abílio Soares, 250 – 12º andar
CEP: 04005-000 – São Paulo (SP), Brazil.
Phone: (11) 3886-4693
FAX: (11) 3886-4695
E-mail: otavioberwanger@yahoo.com.br

INTRODUCTION

Randomized clinical trials are considered the gold standard for scientific studies aiming to assess the effect of a treatment or other interventions (techniques or procedures), during the course of a disease or in a defined clinical situation.

They are quite similar to prospective cohort studies, the difference being their design that permits elimination of some biases such as, selection bias and confusion factors, since treatment and control groups are allocated using randomized techniques and characteristics are distributed similarly in both groups. Furthermore, they are studies submitted to a more intensive control and management.[1]

The idea to distribute a treatment by randomization was proposed by Fisher in 1923, for agricultural research. Successful adaptation of clinical trials to health care in human beings only took place at the end of the forties. The first trial published using a table of randomized numbers for allocation of subjects was the research by Dr. Austin Bradford Hill, of the London School of Hygiene and Tropical Medicine.[2]

Fundamental principles

Chart 1 summarizes the main characteristics of randomized controlled trials:

This is an experimental study in which the single contribution of one factor is isolated, while keeping constant, whenever possible, the other determinants of the outcome. A target population of the intervention is elected, establishing eligibility criteria (inclusion and exclusion). These criteria may be numerous and rigorous when assessing the intervention in a very specific

Chart 1 – Main characteristics of randomized controlled trials

Experimental
Prospective
Controlled
Random and confidential allocation of groups
May or may not have a blinding scheme

clinical situation. In large simple trials these criteria of eligibility are rather brief and simple, to approximate the assessed intervention to clinical practice.[3-4] When these criteria are numerous and rigorous they may limit recruitment of patients and restrict generalization of findings to an overall population. Criteria are planned to increase homogeneity among patients of the study strengthening internal validity.[5]

Randomized allocation permits generation of truly comparable groups so that each patient has the same probability of belonging to one of the groups (exposed or none exposed) as long as all meet the eligibility criteria. All factors related to prognosis and outcome tend to be equally distributed in the comparison groups.[1,2] A such, eventual differences in the occurrence of the outcome among experimental and control groups may always be assigned to the intervention.

Randomization demands special care and must be generated according to an adequate technique. For randomization to be valid, each eligible patient must have an equal chance of being allocated to either one of the study groups without any influence of the researchers. Therefore, it is fundamental that researchers be unable to forecast allocation of the next patients. That is why, allocation of treatment according to the number of the medical chart, date of birth or day of the week are not valid randomization methods.[6]

The list of randomized allocation is usually generated using appropriate software. Other methods considered valid are use of the table of random numbers or even use of dice or coins. Another fundamental item to warrant unpredictability of allocation is to keep the list of randomization confidential, that is to say, researchers must first include the patient in the study and only then will treatment (experimental or control) be defined. This criterion is designated as concealment allocation and is the most important methodological criterion in a randomized controlled trial.[7] The most effective method to assure concealment allocation is central randomization, where researchers register the patient in the study by internet or telephone and later receive the patient's allocation. Another acceptable method is the utilization of sealed envelopes containing the treatment code.

The treatment group may be compared with one or more control groups (arms of the study) that may use the trivial treatment of clinical practice or placebo (in the case of a drug) or even compare efficacy and safety of two different treatments to assess the outcome of interest.

Controlled trials with randomized allocation may refer to drugs, techniques or procedures and may or may not have a blinding scheme.

Blinding takes place together with randomization and means that all those involved in research (participants of the study, researchers, medical team, statistician) are ignorant about the allocation of patients to one group or another. Thus, the study will not be influenced by changes of conduct by the medical team or the patient (Hawthorne effect).[8] Blinding prevents biases at various stages of the research, but cannot always be applied. For instance, assessment of a new surgical procedure in an intensive care unit (ICU) cannot be blind.

CRITICAL EVALUATION OF RANDOMIZED CLINICAL TRIALS

Intensivists interested in reading randomized controlled trials must be alert about several aspects of the study to verify if it has an internal and external validity. Internal validity assesses if the study adequately measured what was proposed, from the point of view of clinical and statistical relevance. External validity analyzes the generalization of results for clinical practices.

Because this is a special type of cohort study, most of the principles utilized for critical assessment of the study are applicable to the randomized controlled trials.[1] Next the most relevant of these will be discussed.

Internal validity of results
For a critical evaluation of the internal validity of a randomized controlled trial, different parameters must be assessed as described below.

Are the exposed and none exposed groups similar except for the exposure factor?
In general, the possible randomization techniques tend to avoid that groups of patients are heterogeneous due to differences attributed solely to chance.[9] However, when a study has an inadequate sample size, homogeneity will not be warranted. Thus, the larger the sample size, the larger the warranty that characteristics and factors related to the study's outcome tend to be equally distributed among

groups.[10] That is why, intensivists should prefer studies that have a sufficient number of subjects, with a satisfactory statistical power, for the assessment of the outcome in relation to the intervention. As such, it is desirable that a comparison table of the baseline characteristics in the treated group and control group be presented to allow assessment of such characteristics. Baseline characteristics refer to the demographic (age, gender) and clinical variables of interest that describe the patients in question. In a study with an adequate sample size these characteristics are rather similar.

If there are differences among the groups, they must be controlled in a statistical analysis.

Was follow-up completed?

Even after randomization some potential sources of loss of follow-up must be taken into account, some patients may not have the disease as believed at beginning of the study, others abandon the research; do not adhere to treatment; present adverse reactions or contact with them is lost. In these cases, comparison between study groups becomes biased, even if design is adequate.[5,11] Suppose a clinical trial where the outcome are thrombotic events, one year after discharge from ICU. While hospitalized, as prophylaxis the treatment group received a new ultra low molecular weight heparin and the control group received another heparin. Now, suppose that when determining the outcome one year later, 30% of the patients who had received this new heparin could not be found. If the analyses of results disclose the non inferiority of this new drug in relation to the control group, these results will have to be cautiously analyzed, because this loss of follow-up may be due to death of many individuals and as such, data were not recorded. When calculations are made again, in truth, treatment does not seem more efficient than then control.

The article must make available the loss to follow-up rates for the critical assessment of results.

Was the study conducted according to the intention-to-treat principle?

This is a methodology which in primary analysis, includes in the results data from all patients in the study and who are analyzed in the arm in which they were initially allocated.[12] It is a principle allowing analysis of results according to the designated treatment and not according to treatment received (explanatory studies). Thus, prognostic factors known or not, will be on the average, distributed equally between groups and all data of the recruited patients will be analyzed, regardless if

the patient has been followed-up until the end of the study. Furthermore, this is why results of the secondary analysis of the study must be carefully assessed, mainly those that exclude data from patients, as there can be loss of randomization effects.[13]

Was the assessed intervention, really the only relevant one in the patient?

When conducting a randomized controlled trial, some situations may mask the true effect of the treatment under study. Often, use of medication together with the utilized product may interfere with its efficacy, due to pharmacodynamic issues, as well as to synergic or antagonistic effects to the medication used. This concept is defined as co-intervention. A way to control such a situation is to include the drug known to interfere, in the exclusion criteria or in the periodical recording of all concomitant drugs administered to the patients during the study for consideration in the estimate of results. If the patient is using some competitive intervention prior to his admission to the study, the influence may be controlled with the utilization of a wash-out period prior to the study.[14]

If the intervention researched is already known and used in clinical practice, the intensivist must ascertain if it is not being used in the control group, out of the study protocol. This situation is defined as contamination in randomized controlled trials, where the proportion of the control group is reduced, since a percentage of it received the intervention. With contamination, the results observed tend to underestimate the real effect of treatment.[14]

Was the study blinded?

Randomized controlled trials using, whenever possible the blinding scheme must be preferred, as already discussed. Measurement of the outcome may be influenced by knowledge of the allocation (observation bias). Patients, as well as researchers may be affected, especially for subjective outcomes such as pain. Additionally, there may be changes in clinical conduct related to knowledge of the allocated. Blinding of patients, healthcare staff and researchers avoids such biases.

Is sample size adequate for the study proposal?

The trial must recruit a sufficient number of patients to demonstrate effectiveness of the intervention. To reach this sample size, various parameters are used.

The sample will be calculated to minimize the two errors α and β. The α or type I error represents the false positive, that is to say, the intervention is not efficient but,

by statistical analysis it is identified as efficient. The α error is minimized by a level of significance chosen, normally of 5%. The β or type II error represents the false negative, that is to say, the intervention is efficient but, by statistical analysis it is identified as not efficient. The type β error is indirectly minimized by the power of the test, since the power is represented by 1 − β. Normally a power of at least 80% is sought, although 90% are hoped for whenever possible. The power of the randomized controlled trial has the ability to demonstrate effectiveness of the intervention.[10]

Consequently, to estimate the sample size some parameters must be considered such as: how will the outcome be measured, as quantitative (numerical) or qualitative (with categories); level of significance desired and power of the test. When the outcome is quantitative it will be necessary to have an idea of the magnitude of the effect and standard deviation. When the outcome is qualitative it will be necessary to know which is the proportion of patients with outcome expected in the intervention group and which is the proportion of patients with outcome expected in the control group.[15]

How should results be presented?

Statistically significant results, that is to say with a probability of significance (p) less than 0.05 may often not be clinically relevant. To assess clinical relevance the estimate of the magnitude of the effect and the accuracy of the estimate must be considered.

How to measure the magnitude of the effect of the intervention?

Measurement of the effects can be made by a ratio, called relative measurement or by a difference, considered the absolute measurement. To measure magnitude of the association (power of the effect) between the intervention and the outcome (how many times the occurrence of the outcome is higher in the intervention group than in control group) we must us relative association measurements. Measurements of association by difference assess how much the frequency of an outcome in the intervention group exceeds that of the control group; that is to say, assess which is the incidence of the outcome attributed to the intervention. The main forms of measuring the effect are summarized below; considering the table 2x2 (Table 1).[15]

- Absolute risk (AR): Is the probability to develop the outcome in each group. It is mathematically represented by a/(a+b) and c/(c+d)

- Absolute risk reduction (ARR): is the difference of ab-

Table 1 – Table 2X2

| Exposure | | Outcome | | |
		Yes	No	Total
	Yes	a	b	a + b
	No	c	d	c + d
		a + c	b + d	

Source: Translated from: Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273(5):408-12.

solute risk between the control group and treated group. It is mathematically represented by ARR = c/(c+d) − a/(a+b). Although reduction of relative risk (described below) is the most utilized parameter for presentation of results, the ARR is the measurement with the greatest clinical importance because it assesses the absolute efficacy of the intervention.

- Relative risk (RR): is the risk of the events between patients in the treated group relative to the risk in the patients of the control group, that is to say RR = [a/(a+b)] / [c/(c+d)]. This measurement tells us the proportion of original risk that is still present when patients receive the experimental treatment.

- Reduction of the relative risk (RRR): Is an estimate of the proportion of baseline risk that is removed by the experimental treatment. There are two ways to estimate it: RRR= 1- RR or RRR = ARR/AR (in the control group). Normally RRR is preferred to RR when presenting results.

- Number needed to treat (NNT): is the measurement used to assess clinical significance. It is mathematically represented by the inverse of the absolute risk reduction. NNT=1/ARR. It expresses the number of patients that must be treated for a time period to obtain a favorable event (in the case of treatment) or to prevent an unfavorable event (in the case of prophylaxis). For instance, if a drug has a NNT equal to five, in relation to the event death, it means that five patients must be treated with it so that one additional death is avoided. Considering a randomized trial in which 50% of the patients die in the control group and 40% die in the treatment group, RRR for death is 20% and NNT to avoid death will be 10 (100/10). This treatment must be preferable to another where NNT to avoid death was 15. But, because different results are possible, a NNT of 10 is not always preferable to an NNT of 15 (if the first was angina and the last was any death, for instance). Therefore, a NNT must always be accompanied by a clearly indicated result and a specified time period, so that diverse interventions can be com-

pared with each other.

- Number Needed to Harm (NNH): is estimated like NNT, however related to absolute risk increase of the intervention.

Supposing a study that wanted to assess a new antibiotic therapy in prophylaxis of hospital infection, consider the results presented in Table 2 below.

Table 2 – Results

		Hospital infection		
New antibiotic therapy		Yes	No	Total
	Yes	5	95	100
	No	15	85	100
		20	180	200

Applying the previously described formula to these data, we have the following interpretations of results:

Absolute risk (AR): Risk of hospital infection in patients who received the new therapy was 5%. For patients who received the standard therapy this risk was 15%.

Absolute risk reduction (ARR): Those who received the new medication had the probability of having hospital infection reduced by 10%.

Relative risk (RR): The risk of events among patients in the intervention group (new therapy) is of 0.33 in relation of the risk in patients of the control group.

It can be stated that the risk of patients of the intervention group is equal to 0.33 or 1/3 of the risk of patients in the control group.

Relative risk reduction (RRR): 67% is the reduction of the risk of having hospital infection in the group of patients who received the new drug in relation to the group who received standard treatment.

Number needed to treat (NNT): in this case, 10 patients must be treated to avoid one case of hospital infection.

Which is the accuracy of the effect of treatment?

Due to the randomized variation, the effect observed in a study, probably will not be exactly similar to that of the "real effect" of the intervention, understanding as real effect the one we would observe in a similar study, however infinitely large. Therefore, a measurement that shows which is the degree of accuracy of the study estimate is necessary, precisely that which is achieved by the confidence interval.[7;16]

The confidence interval of 95%, holds the "real value" of the effect with 95% of probability. The narrower the confidence interval (CI) (close upper and lower limits) the more accurate will be the result. The probability that the true value will stay outside of the interval is of 5 in 100.

If, for instance, we reach a RRR of 25% with CI of 95%: 8% to 40% for a given event, it means that if we repeat the trial 100 times, in 95 of the times the RRR will be found between 8 and 40. Clearly, a RRR of 8% has a different clinical significance from one of 25% or even of 40%.

Although the CI is one of the most important information in the critical assessment of a randomized controlled trial, it is not always mentioned. A study carried out, of analyses of original articles with negative results published in 1997, in the weekly journals British Medical Journal, Journal of American Medical Association, Lancet and New England Journal of Medicine (n=234), had the objective to quantify the proportion of studies with negative results which mentioned the power of the study and its CI.[17] It disclosed that only 30% of the studies mentioned these parameters, while observational studies mentioned them less frequently 15%[95% CI, (8-21%)] than randomized controlled trials 56%, [95% CI,(46-67%),p< 0.001]. They concluded that the prominent medical journals, generally supply insufficient information to assess the validity of the studies with negative results.

All articles should supply the CI, however, when not supplied, we proceed in the following way: 1) should the p value be equal to 0.05 probably the lower limit of the CI for RRR will be zero (we cannot be excluded that treatment is not effective). As the p value decreases, the lower limit of RRR increases. 2) When the article supplies the standard error of RRR (or of RR), the lower and upper limits of CI 95% for one RRR are the points estimated more or less two times the standard error; 3) estimate the CI.

External validation of results

External validation of a randomized controlled trial is related to the study's effectiveness, that is to say the ability to generalize the findings to the whole population subject to receiving the intervention under study. Analyses of external validity involve various aspects such as variations of patient, ethnic cultural variations, factors of severity, cost/benefit considerations, risk, infrastructure, among others. Analysis of the external validity is only justified if analysis of the internal validity of the study was satisfactory. Below the most important considerations are mentioned.

Is there a possibility of generalizing results of the study to the overall population and not only to the study target-population?

Criteria of eligibility of a randomized controlled trial must indicate the population that the researcher intended to infer.

The ideal situation is a study free of biases (internal

validity) which has included patients common to clinical practice (external validity). On the other hand, a study with a high risk of bias, but one that includes a representative sample of the population of interest is of little use. However, we often see studies with rigorous inclusion/exclusion criteria that make recruitment difficult and limit external validity. In these cases the question is, supposing that results are true, can they be applied to my patients?[18,19]

Were the clinically relevant events considered?

Treatment must have an impact on outcomes that are the clinical events of major interest to the patient and to the assisting physician. Clinical outcomes are ranked in five groups: denouement (death), disease (symptoms, physical signs), discomfort (pain, nausea, dyspnea), functional disability (limitation in the capacity to perform common activities) or discontent (emotional reaction to illness or to its care).

Clinical relevance goes beyond statistics and is determined by a clinical trial based on statistical evidence and therefore clear definition of a relevant outcome is needed. Many studies use substitute outcomes that are an indirect measurement (biochemical markers, for instance) or a clinical sign (decrease of the abnormal ventricular depolarization to reduce arrhythmias, for instance) used in substitution of the clinical outcome. The advantage of using substitute outcomes is that the sample size may be minimized because such outcomes usually are more common or because they are continuous variables. Furthermore, the substitute outcome reduces the cost and duration of the study. Disadvantages are that often these measurements, favorable to the intervention, which are first reported, mask the deleterious effects of treatment in other outcomes, perhaps clinically more relevant, that maybe would become manifest only with a longer follow-up.[3,20]

The basis for use of a substitute outcome is that changes produced by the treatment on the substitute outcome must indeed reflect changes in the clinical outcome.

Will benefits of treatment supersede the possible damages and costs?

Frequently, these issues are solved by interim analyses while the study is underway. Assessment and follow-up of any adverse reactions, may often cause interruption of the study before the scheduled date, to assure the patient's safety.[6,21]

Further, a new drug may even prove to be efficient, but if there is another treatment with a more attractive cost/effectiveness ratio, use of this new drug is not justified. The principle "resources are always scarce", used also in

developed countries, must be strictly followed. When it is surmised that resources are always scarce, efficient use of treatments will rely upon allocation of these resources to certain treatments and situations that must display solid evidence of therapeutic efficacy.

CONCLUSIONS

Practice of evidence-based medicine is already a reality when the physician is deciding on the conduct. Nowadays it is not enough to know physiology and pharmacology and to proceed based on our clinical and personal experience to be sure that we are doing our best for the patient.

Critical analysis of evidence requires a profound knowledge on the part of the physician, because evidence-based decision taking is not trivial. Much knowledge is produced daily, but not all published studies have data or quality design, in order to rely upon such information as being true.

Randomized controlled trials are the gold standard of studies for the assessment of interventions, mainly treatment. When well designed and conducted, they provide strong and convincing arguments to support decision taking in clinical practice.

A series of aspects must be considered in the reading of a randomized controlled trial. This article provides an overview of the extent of the knowledge required for the practice of evidence- based medicine.

The intensivist must carefully evaluate the quantity of studies conducted about a clinical issue: and of this sorting, select those with a better internal and external validity and, according to the most convincing and adequate evidence for the profile of his patients, conclude which is the best conduct to be followed.

RESUMO

Ensaios clínicos aleatorizados são investigações científicas consideradas padrão-ouro para avaliar intervenções terapêuticas. Ensaios clínicos aleatorizados podem examinar e avaliar a segurança e eficácia de novas drogas ou procedimentos terapêuticos ou comparar os efeitos entre duas ou mais drogas ou qualquer outra intervenção. Nesse artigo apresentamos as características essenciais e fatores que podem introduzir viés nesses estudos. Em seguida, apresentamos critérios para avaliação crítica de artigos reportando os resultados de ensaios clínicos aleatorizados e mostramos como interpretar e aplicá-los à prática clínica.

Descritores: Avaliação; Ensaios clínicos controlados aleatórios como assunto; Medicina baseada em evidências/métodos; Unidades de terapia intensiva

REFERENCES

1. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. Am J Ophthalmol. 2000;130(5):688.

2. Armitage P. The role of randomization in clinical trials. Stat Med. 1982; 1(4):345-52.

3. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer. 1976;34(6):585-612.

4. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. Br J Cancer. 1977;35(1):1-39.

5. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. Lancet. 2002;359(9308):781-5. Republished in: Z Arztl Fortbild Qualitatssich. 2006;100(6):467-73.

6. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. Lancet. 2002;359(9305):515-9. Comment in: Lancet. 2002;360(9328):258-9; author reply 260. Republished in: Z Arztl Fortbild Qualitatssich. 2007;101(6):419-26.

7. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet. 2002;359(9306):614-8.

8. Schulz KF, Grimes DA. [Epidemiological methods 8: blinded randomized trial: what one covers up is what one obtains]. Z Arztl Fortbild Qualitatssich. 2007;101(9):630-7. German.

9. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: Medical. Stat Med. 1989;8(4):441-54.

10. Schulz KF, Grimes DA. [Failures in sample size calculation in randomized trial: mandatory and mystical]. Z Arztl Fortbild Qualitatssich. 2006;100(2):129-35. Republished from: Lancet. 2005;365(9467):1348-53. German.

11. Sackett DL. Participants in research. BMJ. 2005;330(7501):1164. Comment in: BMJ. 2005;331(7508):109-10. Comment on: BMJ. 2005;330(7501):1175.

12. Lewis JA, Machin D. Intention to treat--who should use ITT? Br J Cancer. 1993;68(4):647-50.

13. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. Lancet. 2002;359(9308):781-5. Republished in: Z Arztl Fortbild Qualitatssich. 2006;100(6):467-73.

14. Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet. 2002;359(9302):248-52. Comment in: Lancet. 2002;360(9328):258.

15. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273(5):408-12. Comment in: JAMA. 2001;286(20):2546-7.

16. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med. 1988;318(26):1728-33.

17. Hebert RS, Wright SM, Dittus RS, Elasy TA. Prominent medical journals often provide insufficient information to assess the validity of studies with negative results. J Negat Results Biomed. 2002;1:1.

18. Sackett DL. Evidence-based medicine and treatment choices. Lancet. 1997;349(9051):570; author reply 572-3. Comment on: Lancet. 1997;349(9045):126-8.

19. Sackett DL. Evidence-based medicine. Spine. 1998;23(10):1085-6.

20. Miller JN, Colditz GA, Mosteller F. How study design affects outcomes in comparisons of therapy. II: Surgical. Stat Med. 1989;8(4):455-66.

21. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. JAMA. 2005;294(17):2203-9. Comment in: Crit Care. 2007;11(1):305. JAMA. 2005;294(17):2228-30.